3.0 STATISTICAL MODELS

This section discusses the statistical models that were fitted to the lead loading, lead concentration, and dust loading data. Also discussed are centering and scaling of design variables to produce easily interpretable model parameters. The stepwise regression and mixed model procedures used to arrive at final models are defined and model parameters are related to specific hypotheses of interest.

Various factors were considered for inclusion in the model. These included abatement and non-abatement factors as fixed effects. To account for within-house and within-room correlation and to estimate house-to-house, and room-to-room variability, random house and room means were included. A discussion of typical levels for the fixed effects, as well as what level was considered as nominal is presented in Section 3.2.

3.1 <u>MIXED RANDOM AND FIXED EFFECTS MODEL</u>

This section describes the statistical models that were fitted to the observed lead loadings, lead concentrations, and dust loadings. These models are the basis for the statistical analyses described in Sections 4 and 5.

The following model contains all of the design factors considered in the study, random effects for house-to-house and room-to-room variation, and additional explanatory variables or covariates. This model was fitted separately to the data for air duct, interior entryway, window channel, and window stool dust samples.

$$\begin{aligned} \ln(C_{ij}) = \ln('') + U_i + R_{ij} &+ \ln(\$_{PI})PI_i + \ln(\$_{PID})PID_i \\ &+ \ln(\$_{SI})SI_i + \ln(\$_{SID})SI_i\,PID_i + \ln(\$_{POD})POD_i \\ &+ \ln(\$_{SO})SO_i + \ln(\$_{SOD})SO_i\,POD_i + \ln(\$_{PR})PR_{ij} \\ &+ \ln(\$_{PRD})PRD_{ij} + \ln(\$_{SR})SR_{ij} + \ln(\$_{SRD})SR_{ij}PRD_{ij} \end{aligned} \qquad [3.1]$$

$+ \ln(\underset{\sim}{(})\underset{\sim}{X}$

for

$i = 1, 2, ... , \#$ houses

$j = 1, 2$ or 3 rooms

where

C_{ij} = measured lead concentration, lead loading, or dust loading in the jth room in the ith house,

" = overall geometric average lead concentration in unabated houses for nominal values of covariates,

U_i = random effect for the ith house; assumed to follow a normal distribution with mean zero and standard deviation F_U,

R_{ij} = random effect for the jth room in the ith house; assumed to follow a normal distribution with mean zero and standard deviation F_R,

$\$_{PI}$ = fixed multiplicative effect associated with a house that has undergone abatement; $\$_{PR}$ is similarly defined for room-level abatement,

PI_i = 1 if abatement was performed in the ith house and zero otherwise; PR_{ij} is similarly defined for room-level abatement,

$\$_{PID}$ = fixed multiplicative effect of interior abatement by E/E methods rather than removal methods; $\$_{POD}$ and $\$_{PRD}$ are similarly defined for outside abatement and room-level abatement,

PID_i = the percentage of interior abatement that was performed by E/E methods; POD_i and PRD_{ij} are similarly defined for exterior abatement and room-level abatement,

$\$_{SI}$ = multiplicative effect of increasing the log-square footage of abatement; $\$_{SO}$ and $\$_{SR}$ are similarly defined for outside abatement and room-level abatement,

SI_i = log-square footage of interior abatement in the ith house or $\ln(1+SFI_i)$ where SFI_i is the square footage of interior abatement in the ith house; SO_i and SR_{ij} are similarly defined for outside abatement and room-level abatement,

$\$_{SID}$ = ratio of the multiplicative effect of increasing the log-square footage of interior abatement by E/E methods to the multiplicative effect of the same increase in the log-square footage of interior abatement by removal methods; $\$_{SOD}$ and $\$_{SRD}$ are similarly defined for outside abatement and room-level abatement.,

$\underset{\sim}{X}$ = vector of additional covariates, and

$\underset{\sim}{(}$ = vector of multiplicative effects associated with increases in the corresponding covariates in the vector X.

The additional explanatory variables (covariates, $\underset{\sim}{X}$) that were considered for inclusion in the model are listed in Appendix B. The variables considered included questionnaire responses, field inspection variables, and measurements taken during the HUD Demonstration. Explanatory variables that were found to be significant for at least one of the sample types are listed by category in the second column of Table 3-1. Nominal values of these covariates and the sample types for which the covariates are significant are listed in the third and fourth columns.

In the model, the " term represents the geometric average lead level that can be expected in houses where no abatement was performed (unabated houses) for nominal values of the covariates included in the model. The random effect term for houses (U_i) allows each housing unit to have its own average lead level. The random effect terms for rooms (R_{ij}) allow each room within the house to have its own average lead level.

The terms PI_i and PID_i and the corresponding coefficients, $\$_{PI}$ and $\$_{PID}$, allow estimation of the effect of abatement and also allow a distinction between the effects of different abatement methods. $\$_{PI}$ characterizes the abatement effect without distinguishing between E/E methods and removal methods. $\$_{PID}$ characterizes the difference in the interior abatement effects

Table 3-1. Explanatory Variables that are Significant for at Least One Sample Type

Explanatory Variable Category	Explanatory Variable	Nominal Value	Sample Types for Which Explanatory Variable is Significant
Abatement	Abatement contractor	Average across contractors	ARD
	Total Interior Abatement	282 for Typical E/E 61 for Typical Removal 180 for Typical Abated	FLR, WCH, WST
	Total Exterior Abatement	628 for Typical E/E 260 for Typical Removal 519 for Typical Abated	WCH, FDN
	Phase of HUD Demonstration (of three) in which residence was abated	Average across phases	WST
	HUD XRF or AAS measure of paint lead loading	Control: 0.10 (mg/cm^2) Abated: 0.44 (mg/cm^2)	FDN
	Specific removal method used in a room – chemical stripping – remove and replace – heat gun – removal	 15% 15% 30% 40%	WCH
Substrate	Substrate type	Average across substrates	FLR, EWI, WCH, FLW
	Substrate condition	Good	ARD, WCH
Cleanliness	Frequency of wet mopping uncarpeted floors	12/month	ARD
	Frequency of window stool dusting	1/month	ARD
	Frequency of vacuuming uncarpeted floors	12/month	EWI, EWO, FLR
Occupation	Wearing home work clothes from an occupation with potential lead contamination	No	WST, EWY
	Resident employed in welding occupation	No	FDN, FLR
	Resident employed in salvage occupation	No	BDY
	Resident employed in paint removal occupation	No	BDY
Activities	Frequency of removing paint at home	Not in last 6 months	EWI, FDN
	Frequency of pipe or electrical component soldering	Not in last 6 months	BDY
Other resident factors	Year house was built	Control: 1943 Abated: 1926	BDY, FDN, EWY
	Number of children (between ages of 7 and 17)	0	EWI
	Months at residence	18	FDN
	Ownership of home	Owner	FDN
	Number of pets	0	FLR
Sampling deviations	Air duct samples taken from cover of air duct	No	ARD
	Window channel samples taken with small nozzle	No	WCH

for E/E methods versus removal methods. Exterior and room-level abatement effects are handled similarly in the model.

The term SI_i and the corresponding coefficients, $\$_{SI}$ and $\$_{SID}$, allow the effect of the amount of interior abatement, on a per log-square foot abated basis, to be estimated by the model. $\$_{SI}$ characterizes the interior abatement effect per log-square foot abated without distinguishing between E/E methods and removal methods. $\$_{SID}$ characterizes the difference in the interior abatement effects per log-square foot abated for E/E methods versus removal methods. Exterior and room-level abatement effects are handled similarly in the model.

In the case of floor dust vacuum samples, an additional within-room random error term was added to model [3.1],

$,_{ijk}$ = random effect for the kth sample in the jth room of the ith house.

Floor dust wipe samples were taken from only one location in each of the abated houses. Therefore, no room level effects were included in the model, nor can differences between abated and unabated houses be estimated. The following model was used for these samples:

$$\begin{aligned} \ln(C_{ij}) = \ln('') + U_i + R_{ij} &+ \ln(\$_{PID})PID_i \\ &+ \ln(\$_{SI})SI_i + \ln(\$_{SID})SI_i\,PID_i + \ln(\$_{POD})POD_i \\ &+ \ln(\$_{SO})SO_i + \ln(\$_{SOD})SO_i\,POD_i \\ &+ \ln(\underset{\sim}{(})\underset{\sim}{X}. \end{aligned} \qquad [3.2]$$

The model fitted to the data for exterior entryway dust samples is

$$\begin{aligned} \ln(C_{ij}) = \ln('') + U_i + S_{ij} &+ \ln(\$_{PI})PI_i + \ln(\$_{POD})POD_i \\ &+ \ln(\$_{SO})SO_i + \ln(\$_{SOD})SO_i\,POD_i + \ln(\underset{\sim}{(})\underset{\sim}{X} \end{aligned} \qquad [3.3]$$

where

C_{ij} = measured lead concentration at ith house,

S_{ij} = random effect for the jth side of ith house; assumed to follow a normal distribution with mean zero and standard deviation F_S,

and all other terms are defined as above. For exterior samples, the random side effect, S_{ij} takes the place of the random room effect, R_{ij} .

For foundation soil, boundary soil, and entryway soil, an additional within-side of house component of variation is added to model [3.3]:

E_{ijk} = random effect for the kth sample on the jth side of ith house; assumed to follow a normal distribution with mean zero and standard deviation F_E,

The third objective of this study was to investigate the relationships between lead in household dust and lead from other sources. The estimated house-level and room/side-level random effects for the different sample types provide a basis for this investigation. A discussion of these relationships is provided in Section 5.0.

3.2 CENTERING AND SCALING OF COVARIATES

Several covariates included in the models were centered and scaled so that the model parameters would have more meaningful interpretations. In order to determine the appropriate centering and scaling parameters, three classes of abated houses were identified: (1) predominantly E/E, (2) predominantly removal, and (3) abated. The third class is the combination of the first and second classes. As illustrated above in Tables 2-3 and 2-4, a different combination of E/E and removal methods was applied in each house. Each house was classified separately for interior and exterior abatement. For interior sample types, if the percentage of interior abatement performed by E/E methods was more than 50%, then the house was classified as predominantly E/E. Otherwise, it was classified as predominantly removal. A similar approach was used for exterior sample types.

For each of the three classes of abated houses two quantities were determined:

- Typical percentage abated by E/E methods, and
- Typical square footage abated.

These values are reported in Table 3-2 for interior, exterior, and room-level abatement. The typical percentage abated by E/E methods was determined by taking an average across all houses in the class.

A correlation was observed between total square feet abated in a house and the method used to perform the abatement. Typically, significantly more square feet were abated when E/E methods were used than when removal methods were used. This occurred both indoors and outdoors. Therefore, the typical square footage abated was treated as a function and allowed to vary with the percentage abated by E/E methods. To accomplish this, a simple linear regression of log-square feet abated versus percent abated by E/E methods was fitted to the data for all abated houses. Figure 3-1 displays the regression relationship for interior abatement. Similar regression relationships were developed for exterior and room level abatement.

The typical square footage abated values reported in Table 3-2 are taken from the regression relationship for the typical percentage abated by E/E methods. Taking interior abatement for example, a predominantly E/E house with 93% E/E abatement is predicted to have 282 total square feet of interior abatement.

Similarly, a predominantly removal house with 4% E/E abatement is predicted to have only 61 total square feet of interior abatement. Finally, an abated house with 67% E/E abatement

Table 3-2. Average Percent Abated by E/E Methods, by Abatement Method Classification for Interior, Exterior and Room Level Abatement

Level	Typical % Abated by E/E Methods			Typical Square Footage Abated		
	E/E	Removal	Abated	E/E	Removal	Abated
Interior	93	4	67	282	61	180
Exterior	92	27	78	628	260	519
Room	96	3	69	70	36	58

Table 3-3. Centering and Scaling Parameters for Model Covariates

Covariates	Value Subtracted		Value Divided By
	Control	Abated	
PID	0	67%	89%
POD	0	78%	65%
PRD	0	69%	93%
SI	0	ln(57)+0.0172*(E/E%)	ln(2)
SO	0	ln(180)+0.0136*(E/E%)	ln(2)
SR	0	ln(35)+0.0072*(E/E%)	ln(2)
PR	0	1	-1

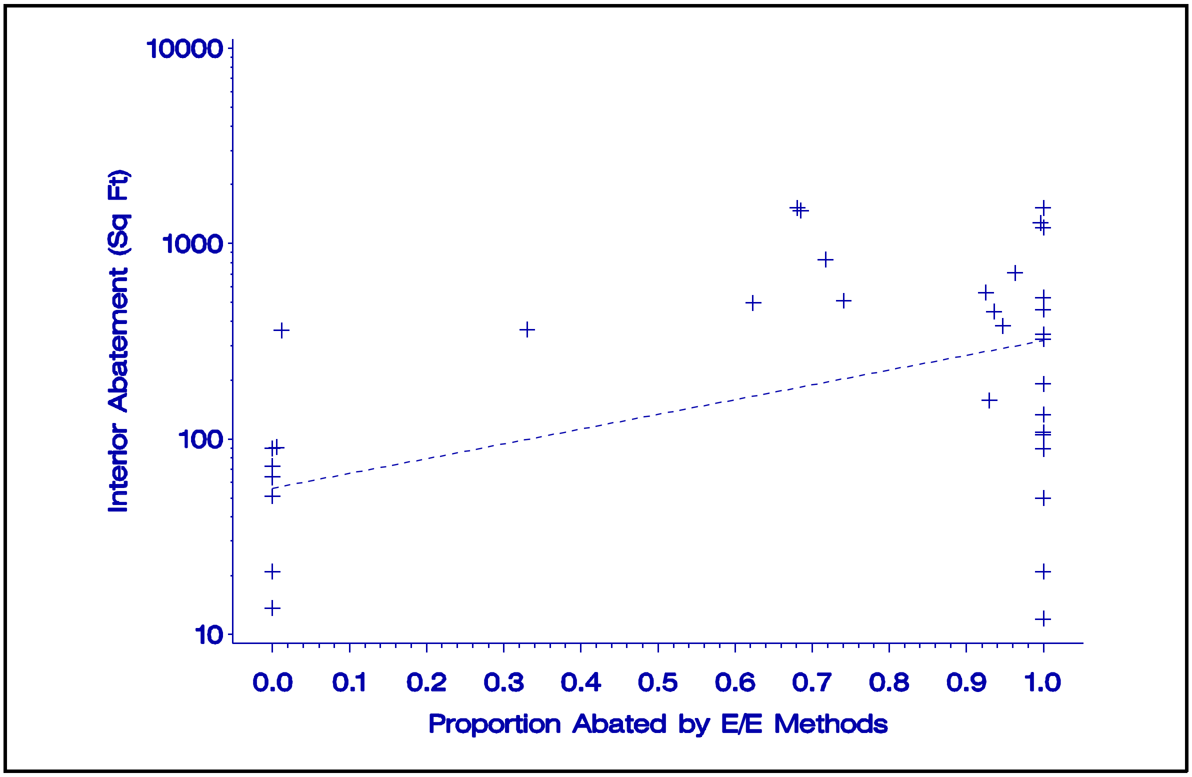

Figure 3-1.Total square feet abated indoors vs. percent encapsulated/enclosed indoors: Abated units.

is predicted to have 180 total square feet of interior abatement. The typical square footage abated values in Table 3-2 for exterior and room level abatement were determined in a similar fashion.

Table 3-3 describes how the values presented in Table 3-2 were used to center and scale the model covariates so that the model estimates have a meaningful interpretation. Table 3-4 displays the interpretation of each of these factor effects after transformation. These interpretations are consistent with the hypotheses we wish to test, as will be discussed in Section 3.4. For abated houses, PID, POD, and PRD values were centered by subtracting off the typical percent abated by E/E methods for an "abated" house. These values were then scaled by dividing the

Table 3-4. Parameter Interpretation After Centering and Scaling

Parameter	Interpretation
$\$_{PI}$	Ratio of the expected lead level in a typical abated room in a typical abated unit[a] to the expected lead level in a control unit
$\$_{PO}$	Ratio of the expected soil lead level for a typical abated unit[a] to the expected soil lead level for a control unit
$\$_{PR}$	Ratio of the expected lead level in a control room in a typical abated unit[a] to the expected lead level in a typical abated room in the same abated unit
$\$_{PID}$	Ratio of the expected lead level in a typical abated room in a typical E/E unit[a] to the expected lead level in a typical abated room in a typical removal unit
$\$_{POD}$	Ratio of expected soil lead level for typical E/E unit[a] to expected soil lead level for typical removal unit
$\$_{PRD}$	Ratio of the expected lead level in a typical E/E room in an abated unit to the expected lead level in a typical removal room in the same abated unit
$\$_{SI}$	Multiplicative effect of doubling the square footage of interior abatement in a typical abated unit[b]
$\$_{SO}$	Multiplicative effect of doubling the square footage of exterior abatement in a typical abated unit[b]
$\$_{SR}$	Multiplicative effect of doubling the square footage of room-level abatement in a typical abated room[b] while holding the house total square footage constant and mix of unit level abatement constant
$\$_{SID}$	Ratio of the multiplicative effect of doubling the square footage of interior abatement in a typical E/E unit[b] to the multiplicative effect of doubling the square footage of interior abatement in a typical removal unit[b]
$\$_{SOD}$	Ratio of the multiplicative effect of doubling the square footage of exterior abatement in a typical E/E unit[b] to the multiplicative effect of doubling the square footage of exterior abatement in a typical removal unit[b]
$\$_{SRD}$	Ratio of the multiplicative effect of doubling the square footage of room-level abatement in a typical E/E room[b] to the multiplicative effect of doubling the square footage of room-level abatement in a typical removal room[b] while holding the house total square footage constant and mix of unit level abatement constant

(a) Typical with respect to both E/E% and square footage abated as indicated in Table 3-2.

(b) Typical with respect to E/E% as indicated in Table 3-2 but with varying square footage abated.

centered variable by the difference between the typical percent abated by E/E methods for a typical "E/E" house minus a typical "removal" house. For example, to obtain the variable PID, 0.67 was subtracted from the percent of interior abatement performed by E/E methods, and then this difference was divided by 0.89 (= 0.93 - 0.04). The result is a variable whose effect can be interpreted as the following ratio:

$$\frac{\text{Expected lead level in a typical abated room in a typical E/E house}}{\text{Expected lead level in a typical abated room in a typical removal house}} .$$

SI, SO, and SR values were centered by subtracting off the logarithm of the predicted square footage abated based on the regressions versus E/E percentage discussed above. These values were then scaled by dividing by ln(2). Finally for abated houses, PR (the unabated room indicator) was subtracted from one (making abated rooms the default for abated houses). The values of these variables in unabated houses were left as zero.

Information on many of the factors determined to be significant was obtained during an interview with a resident of each house sampled. A summary of the interview results is provided in Appendix E. Before models were fitted, these factors were also centered at nominal levels. Centering was accomplished by subtracting off the nominal value reported in Table 3-1. Some factors, such as age of home and XRF measures were very correlated with the abatement indicator. In these cases a nominal level was determined both for the unabated houses and for the abated houses. The estimated effect then represents the effect of the factor above and beyond the effect of abatement. These nominal levels are reported again in Section 4 in each table where estimates are given, along with the scaling factor used. The selection of nominal values is also discussed in more detail in Section 4.

The purpose of including XRF measures as a covariate was to control for differences in pre-abatement lead levels. In rooms where XRF measures were taken during the HUD Demonstration, a geometric average was calculated. However, due to the variability in observed XRF levels, negative values were obtained in several cases. Since it is impossible to have a negative amount of lead and the smallest positive reading by the XRF was 0.1, these values were regarded as censored at 0.1 mg/cm^2, and a censored mean for the room was estimated. If only one component was measured within a room, and the reading was at or below 0 mg/cm^2, 0.05 mg/cm^2 was used in the analysis; if more than one component was measured and all were reported at or below 0 mg/cm^2, 0.07 mg/cm^2 was used.

3.3 MODEL SELECTION

The procedure used to select models to fit to the data was developed in concert with the study objectives. Specific terms corresponding to the primary design factors were included in the model to test hypotheses associated with the objectives of the study. These hypotheses are listed in Section 3.4.

Every model used in this study included the following primary design factors:

- A term to distinguish between unabated houses and abated houses (PI), and
- A term to distinguish between abatement methods (PID for interior samples, POD for exterior samples).

Models for interior dust measurements also contained:

- A term to distinguish between unabated rooms and abated rooms in abated houses (PR).

There is one exception. All wipe floor samples were taken in only one room of abated houses. Although for 4 of the 34 houses these samples were collected from a unabated room, room-level abatement effects were not estimated from the data collected by wipe sampling.

In addition to the three primary design factors, many additional factors (questionnaire data, field observations) were included to estimate other effects which may affect lead levels. The additional factors included in each model were selected using a phased stepwise regression approach.

3.3.1 Phase 1: Abatement Effects (Stepwise Regression)

First, stepwise regression was used to select additional abatement design factors which were significant above and beyond the effects of the three primary design factors described above. The additional abatement factors considered included square-footage abated by room, as well as a breakdown of square-footage by abatement method.

In the stepwise regression, factors were retained only if they were significant at the 5 percent level. Any factor found to be significantly associated with either lead concentration or lead loading was automatically forced to be retained in the model for the next selection phase.

3.3.2 Phase 2: Non-Abatement Factors (Stepwise Regression)

In a second phase of factor selection, all remaining factors were considered as candidate factors in addition to the design factors discussed above. These included questionnaire and visual observation data, HUD Demonstration Data, and other practical measures. Appendix B presents a list of all the factors considered for inclusion in the models. Stepwise regression was used again to select significant factors. Any factors found to be significant at the 5 percent level were retained for the next selection phase.

To avoid confounding, a preliminary correlation analysis was performed to screen any factors which were strongly correlated with others. For example, for 15 of the 16 homes in which a resident wore work clothes home from their occupation, their clothes were also washed at home. Therefore, only the former was included as a candidate factor in the stepwise regression. Specifically, if any factor was more than 80 percent correlated with another, one of the factors was excluded from the models. The factor with the most complete data was used in fitting the models.

3.3.3 Phase 3: Mixed Model Screening (Backward Elimination)

Phase 1 and Phase 2 identified a subset of factors with some association with lead levels. However, due to software limitations, the stepwise regressions were based on fixed effect models whereas it is proper to use a mixed model with random effects in the factor selection process described above. Therefore a mixed model was fitted with random house and random room/side of house effects where appropriate. Any factors not found to be significant by the mixed model analysis at the 10% level were removed from the model (aside from the three design factors described at the beginning of Section 3.3). This process was repeated, refitting the model each time and removing one factor at a time, until all factors remaining were observed as significant covariates for either lead loading or lead concentration.

The final models varied by sample type. Appendix C displays the selected factors and their estimated effects by sample type and response (lead concentration, dust loading, lead loading). This table is explained in more detail in Section 4.

3.4 <u>HYPOTHESIS TESTS</u>

Data were collected to test the following hypotheses:

H_{01}: Average lead levels in a typical abated room in a typical abated house are equivalent to average lead levels in an unabated house.

H_{02}: Average lead levels in a typical abated room in a typical E/E house are equivalent to average lead levels in a typical abated room in a typical removal house.

H_{03}: Average lead levels in a typical abated room in a typical abated house are equivalent to average lead levels in a unabated room in a typical abated house.

H_{04}: House to house differences above and beyond those explained by the models are uncorrelated.

Hypothesis H_{01} is equivalent to the hypothesis that $\$_{PI}=0$, hypothesis H_{02} is equivalent to the hypothesis that $\$_{PID}=0$, and hypothesis H_{03} is equivalent to the hypothesis that $\$_{PR}=0$. Thus, the model parameters align perfectly with the hypotheses to be tested. Hypothesis H_{04} will be tested via extensive correlation analyses in Section 5.